# Grouped Goodness-of-Fit Tests for Binary Regression Models

by

**Jana Dorthea Canary**

**BS (Mathematics), MS (Forest Ecology)**

**Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (Biostatistics)**

**University of Tasmania, December 2013**

| | |
|---|---|
| **Supervisors** | **Doctor Stephen Quinn** |
| | **Associate Professor Leigh Blizzard** |
| | **Professor David Hosmer** |
| **Research Supervisor** | **Professor Ronald Barry** |

# Dedication

To my parents Willa and Jim, my husband Mark, and my daughter Jacqueline

# Declaration of Originality

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signed: ………………………………………. Date: ……………………………

## Statement of Authorship

This thesis can be made available for loan. Copying of any part of this thesis is prohibited for two years from the date this statement was signed; after that time limited copying is permitted in accordance with the *Copyright Act 1968.*

Signed: ………………………………………. Date: ……………………………

# Acknowledgements

I would like to first thank my primary supervisor, Dr. Stephen Quinn, for his support. I feel fortunate to have had Steve as a supervisor. He was always willing to take time to work with me, to give carefully thought out comments on my work, and to deal with the challenges of working with someone on the other side of the globe. I am very appreciative of his efforts in guiding me through the thesis process.

I would also like to thank my other supervisors. First, thank you to Dr. Leigh Blizzard, who made it possible for me to come to Menzies and helped me to navigate through the steps of setting up a Ph.D. candidature remotely from Alaska. He also provided useful comments on my work and provided financial support through an NHMRC grant that, along with an Australian Postgraduate Award, made my research possible. I also would like to thank Dr. Ron Barry of the University of Alaska, who was kind enough to meet with me to discuss my work while I was in Alaska, particularly during the early stages of my candidature. Lastly, I would like to thank Dr. David Hosmer, who is an emeritus professor and an expert in the field of biostatistics. He kindly offered his time and advice, and I very much appreciate his help.

I would also like to thank several fellow students and staff (in no particular order) at Menzies whose support both in person and electronically made the process so much easier: Kara, Laura, Dawn, Peta, Oliver, Kathy, Karen, Petr, Barbara, Tracey, Steve, David, and Ben. Thank you for your advice, support, encouragement, and friendship. Also my friends in Alaska and Hobart who provided childcare when I needed it, chats, and support: Petra, John, Gordon, Nancy, Jay, Donie, Peter, Margaret, Trusten, William, Leslie, and Sharon.

Also, I would like to thank the two anonymous examiners of my thesis. Their thorough review and thoughtful comments were very helpful and encouraging.

Lastly I would like to thank my husband Mark Conde and daughter Jacqueline. I cannot thank Mark enough for his financial and emotional support, and for lively discussions about mathematics. And thank you to Jacqueline, for the backrubs, hugs, and understanding when I did not have time to do fun things. I hope to make it up to you soon.

# Abstract

How well a proposed regression model fits the observed outcome data is a critical question. The answer may influence model selection, and the conclusions drawn. Summary goodness-of-fit (GOF) statistics are used to assess model fit. Pearson's chi-squared GOF statistic $\left(X^2\right)$ is used to evaluate the fit of logistic regression models, but $X^2$ isn't appropriate when the model contains continuous covariates. Other GOF statistics are applicable, including the Hosmer-Lemeshow $(HL)$, Pigeon-Heyse $\left(J^2\right)$, and Tsiatis $(T)$ statistics. All have similarities to $X^2$ and group data artificially.

Simulation studies assessing new GOF statistics for logistic models with continuous covariates often include $HL$ for comparison. We know of no study that compares $HL$, $J^2$, and $T$. We did so here, applying the same grouping method (deciles-of-risk) to all. Our results indicated that $HL$ and $T$ followed their reported distributions, but $J^2$ did not. Its distribution was closer to $J^2 \sim \chi^2(G-2)$, where $G$=groups, rather than the reported $\chi^2(G-1)$. Assuming $J^2 \sim \chi^2(G-2)$, $T$ maintained the Type I error rate twice as often as $HL$ and $J^2$. The rates of $HL$ and $J^2$ were often lower than expected when dichotomous, quadratic, or interaction terms were included. The statistics had similar power to detect departures from a true underlying model.

The logistic model is the canonical generalized linear model (GLM) for binomial outcomes. Although many GOF statistics have been developed for logistic models, there are fewer for non-canonical GLM with binomial outcomes. The properties of the logistic model make the development of GOF statistics relatively straightforward, but it can be more difficult for non-canonical GLMs.

We considered whether $HL$, $J^2$, and $T$ could be applied to non-canonical GLM with Bernoulli outcomes and continuous covariates. Our investigation found that $HL$ and $J^2$ can be applied directly, but $T$ cannot. We introduced an augmented version of the Tsiatis model and

generalised $T$, ($T_{\mathcal{G}}$). We showed that under non-canonical links, $T_{\mathcal{G}} \sim \chi^2(G)$. In a second simulation study, $HL$, $J^2$, and $T_{\mathcal{G}}$ were used to evaluate the fit of probit, log-log, complementary log-log and log binomial models. The deciles-of-risk method was applied. Type I error rates were consistently maintained by $T_{\mathcal{G}}$, while those of $HL$ and $J^2$ were often lower than expected if the model included dichotomous, quadratic, or interaction terms. Because the distributions of $HL$ and $J^2$ varied, it was unclear how their degrees-of-freedom could be adjusted. The statistics had similar power to detect an incorrect model in most situations. An exception occurred when a log model was incorrectly fit to data generated from a logistic model; here $T_{\mathcal{G}}$ had more power than $HL$ or $J^2$.

# Table of Contents